

백서

비즈니스 속도에 따른 하이브리드 분석 제공: IBM Integrated Analytics System(IAS)

후원사: IBM

Carl W. Olofson
October 2018

백서 내용

본 백서에서는 관련된 데이터의 양과 다양성 및 속도가 증가함에 따라 기업이 직면하게 되는 부담과 데이터를 이해하는 데 소요되는 시간이 수용할 수 없을 정도로 오래 걸리는 데에 주목하고 있습니다. 의사 결정을 내릴 수 있는 분석을 위해 통계 데이터를 유용한 방식으로 활용하려고 하는 대부분의 IT 환경에서는 여러 가지 출처의 데이터를 수집하여 특별한 환경 설정과 조정이 필요한 관계형 데이터베이스에 모아 놓아야 데이터 전문가를 통해 비즈니스 분석가에게 유용한 모델을 구축할 수 있습니다. 그리고 이로 인해 발생하는 복잡성은 프라이빗 및 퍼블릭 클라우드 시스템은 물론이고 온프레미스(on-premises) 형태의 전통적인 데이터 센터에 있는 데이터를 수집하고 분석해야 할 필요성에 따라 더욱 가중됩니다. 이는 환경 설정에 있어 하이브리드 클라우드 환경 지원을 요구하고 있습니다. 본 백서에서는 이러한 문제를 설명한 이후 관련 데이터에 대한 접근 및 관리를 촉진하고 쿼리를 통한 강력한 분석과 신속한 의사 결정이 이어질 수 있을 만큼 높은 성능으로 설계된 목적 기반 데이터베이스 시스템의 유용성을 고려합니다. 또한 IBM Integrated Analytics System(IAS)가 온프레미스, 퍼블릭 클라우드 또는 프라이빗 클라우드에서 이러한 하이브리드 기능을 어떻게 제공하는지 살펴 봅니다.

상황 개요

기업은 운영 상태를 나타내는 광범위하고 관리되지 않은 출처로부터 오는 데이터 혹은 마케팅이나 운영상의 필요와 관련된 외부 데이터를 채택하려는 움직임을 접하게 될 때 새로운 도전 과제에 직면하게 됩니다. 데이터를 수집하고 유용한 구조로 변환한 다음에 데이터를 사용하여 인간과 주로 머신러닝(ML, Machine Learning)에 의해 주도되는 기계 양쪽 모두에 기반한 의사 결정을 통해 통찰력과 운영을 이끌어 내야 합니다. 그 결과 소위 디지털 트랜스포메이션을 통해 비즈니스 관리가 완전히 변경되고 있고, 주문형 확장 가능 리소스에 대한 요구 사항으로 인해 많은 기업이 클라우드로 전환하고 있습니다.

데이터 수집, 큐레이션 및 통합의 기술적 문제

기업은 디지털 트랜스포메이션 과제를 어떻게 해결할 수 있는지 고려하면서 예기치 못한 범위의 다양한 애플리케이션과 사용 사례 및 비즈니스 요구 사항을 처리할 수 있는 기술을 찾아야 합니다. 이러한 기술은 외부 스트리밍 데이터, 사물 인터넷(IoT) 데이터, 그리고 내부에서 수집한 데이터를 포함한 다양한 출처와 형식 및 유형의 데이터를 수집할 수 있어야 합니다. 또한 데이터를 정의하고 정리 및 정렬하여 함께 가져올 수 있어야 하며, 이후에 원시 데이터에서 상세한 분석, 기계 학습, 그리고 비즈니스 계획에 사용될 수 있는 특정 모델 데이터를 수집해야 합니다.

이 기술에는 엔터프라이즈 엣지와 데이터 수집과 조직 및 초기 분석에 사용되는 데이터 레이크를 의미하는 참여된 시스템에서 데이터를 관리하는 데 사용되는 NoSQL 시스템과 비즈니스 분석 및 의사 결정 지원을 위한 관계형 데이터 웨어하우스가 포함될 수 있습니다. 또한 오픈 소스 커뮤니티를 포함하여 다양한 공급 업체에서 제공하는 오픈 소스 기반의 제품 및 라이선스가 부여된 소프트웨어가 포함됩니다. 소프트웨어 기술의 요구 사항은 다양할 수 있으며, 데이터 저장과 처리를 최적화할 수 있도록 유연하게 배치된 서버 클러스터가 요구될 수 있습니다.

시간이 지남에 따라 이러한 시스템과 운영을 관리, 조정 및 통합하는 것은 매우 어려울 수 있습니다. 소프트웨어를 구입 및 업그레이드 하고 패치를 적용할 때 버전 호환성을 추적하는 데 필요한 광범위한 기술과 관리가 요구됩니다. 이러한 다양한 기술을 다양한 출처에서 얻어 수동으로 관리할 경우 비용 초과 및 프로젝트 실패가 발생할 수 있습니다.

관련된 요소에는 다음과 같은 결합이 포함되기도 합니다.

- 일반적으로 하둡(Hadoop) 기반의 데이터 레이크 관리 플랫폼
- 데이터 필터링, 정렬, 중복 제거, 정의 및 보안 기술
- 스트리밍 데이터 처리 및 라우팅 엔진
- 스파크(Spark)와 같은 고성능의 일반화된 데이터 분석 플랫폼
- 선택된 데이터를 데이터 레이크 또는 스트림 저장소에서 엔터프라이즈 데이터 웨어하우스로 이동할 수 있는 데이터 통합 기술
- 확장 가능한 고성능 관계형 데이터베이스 관리 시스템으로 관리되는 데이터 웨어하우스
- 데이터의 쿼리, 검색 및 시각화를 위한 분석 도구
- 스마트 시스템에 권한을 부여하는 데 사용할 수 있는 데이터 기반의 인공 지능 및 ML 소프트웨어
- 상기에 언급한 모든 요소를 지원할 수 있게 잘 구성된 서버, 저장소 및 네트워킹

시스템으로 관리하기

위와 같은 요소를 하나의 시스템으로 통합하고 관리하기 위해서는 상당한 작업의 프로젝트가 필요합니다. 기술을 신중하게 선택하고 이를 통합하기 위해서는 아키텍처를 개발해야 하고 비즈니스 요구 사항이 변하면 이에 따라 시스템의 모든 요소를 관리, 조정 및 조율하면서 이를 향상시키는 프로세스가 필요합니다. 이러한 프로젝트는 1년 넘게 지속될 수 있지만 최종적으로 성공을 보장하지는 못합니다.

이상적으로는 단일 공급 업체가 하나의 제품으로 제공하는 잘 설계된 기술 조합을 찾는 것이 좀 더 괜찮은 계획입니다. 해당 제품은 잘 통합된 모든 기능 요소를 포함해야 할 뿐만 아니라 간단하고 직접적인 설계 및 관리 기능으로 한꺼번에 실행할 수 있는 하드웨어 플랫폼을 제공해야 합니다. IBM Integrated Analytics System 이 하나의 예가 될 수 있습니다.

IBM Integrated Analytics System (IIAS)

IBM Integrated Analytics System 은 여기에서 설명된 모든 요소를 하나로 묶어 단일 시스템으로 통합합니다. 지금까지 IBM 은 엔터프라이즈 분석용 퓨어데이터 시스템(PureData System)과 운영 분석용 퓨어데이터 시스템 (PureData System) 두 가지 주요 분석 시스템을 제공하고 있습니다. IBM Integrated Analytics System 은 BLU 액셀러레이션(Acceleration)을 포함하여 상기 두 시스템의 기능을 통합 솔루션으로 결합하고 있습니다. 네티자(Netezza) 기술 개발에 20년 이상 역점을 두어 운영 분석용 퓨어데이터 시스템(이전에는 Netezza 로 알려짐)을 주도한 차세대 기술이 핵심입니다.

모든 구조화된 데이터 분석 컨텍스트에서 사용하기 위한 것으로 매우 짧은 대기 시간의 데이터 분석에 적용할 수 있습니다. 즉, 생성 후 몇 분 내에 데이터를 수집하여 처리합니다. 또한 휴대 전화 통화 데이터 레코드(CDR)와 같이 대용량의 기계 생성 데이터를 처리할 때에도 1 초 미만의 응답 시간으로 처리하도록 설계되었습니다. 높은 동시성을 지원하면서도 상기와 같이 작동하도록 설계되어 수천 명의 사용자가 대시 보드 및 경고와 같은 메커니즘을 통해 데이터를 원하는 대로 편집하고 이를 통해 적시에 데이터를 볼 수 있습니다. 스파크의 임베딩과 파이썬 (Python) 및 R 에 대한 지원을 통해 IBM Integrated Analytics System 은 IBM DSX(Data Science Experience)가 제공하는 일류의 데이터사이언스 기능을 제공함으로써 고객이 데이터 사이언스와 기계 학습을 대규모로 수행할 수 있도록 지원합니다.

IBM Integrated Analytics System 의 목적은 관계형 데이터베이스와 NoSQL 데이터베이스 기술을 통합하고, 구조화된 데이터 및 구조화되지 않은 데이터를 모두 지원하는 온프레미스와 클라우드 또는 하이브리드 클라우드 구성에서 배포를 지원하는 것입니다. IBM Integrated Analytics System 은 커먼 SQL 엔진(Common SQL Engine)으로 고정된 클라우드와 온프레미스 구조형 및 반구조화된 데이터 플랫폼 전반에 완벽한 쿼리를 지원합니다.

IBM 커먼 SQL 엔진은 유연성, 애플리케이션 호환성과 이식성, 강력한 데이터 통합, 가상화 및 완벽한 쿼리를 온프레미스 또는 클라우드에서 제공하는 IIAS 의 핵심 요소입니다. 일반적으로 데이터 배포의 경우 다양한 데이터를 사용하는 쿼리 및 애플리케이션/관리 체계를 다시 작성하거나 구조를 조정해야 합니다. 커먼 SQL 엔진에는 오라클 애플리케이션 호환성 계층(Oracle Application Compatibility Layer)이 포함되어 있어 오라클 애플리케이션이 IIAS 뿐만 아니라 IBM Db2 제품군과도 통합될 수 있습니다. 일반적으로 기존 오라클 애플리케이션 코드의 98% 이상은 그대로 실행될 수 있습니다. 일부 수작업은 IBM 에서 말한 것처럼 "데이터베이스 배포나 공급 업체 또는 형식을 변경할 때 혹은 애플리케이션을 다시 작성하거나 새 소프트웨어를 라이선스 처리할 때, 또는 두 가지 모두를 수행할 때" 요구되기도 합니다.

IIAS 는 IBM 커먼 SQL 엔진과 함께 IBM Db2 애널리틱스 액셀러레이터(IBM Db2 Analytics Accelerator, IDAA)를 활용하여 쿼리를 신속하게 실행하고 핵심 정보가 필요한 사용자와 애플리케이션으로 전달합니다. IDAA 와 IBM 커먼 SQL 엔진은 복잡한 비즈니스 분석을 위해 독보적인 혼합 워크로드 성능을 제공하기 때문에 다음과 같은 기능을 수행할 수 있습니다.

- 대량의 데이터를 신속하게 분석
- 분석, 인덱스 디자인, 쿼리 재작성 등의 DBA 및 데이터 디자인 작업 불필요
- 복잡한 쿼리 작업 부담을 줄임으로써 월간 라이선스 비용 절감

IIAS 는 시스템 어디서나 기록된 데이터에 쉽게 접근할 수 있습니다. 이 시스템의 요소에는 다음 사항이 포함됩니다.

- 관리되는 퍼블릭 클라우드 DBaaS
- 온프레미스 또는 클라우드에서 소프트웨어 정의 웨어하우스
- 분석 전용 어플라이언스
- HDFS, NoSQL 데이터베이스, 객체 저장소 및 WebHDFS 에 대한 접근을 통합할 수 있는 BigSQL 을 사용하여 하둡 (Hadoop)과 연합하고 쿼리를 데이터로 이동할 수 있는 기능
- 사용자 정의 배포 가능한 데이터베이스
- 호튼웍스(Hortonworks)가 지원하는 오픈 소스 하둡(Hadoop)

이 시스템은 고성능의 최적화된 클라우드 지원 데이터 플랫폼으로 데이터 전문가를 위해 개발과 배포 시간을 단축할 수 있게 설계되었습니다. 또한 z 시스템의 애플리케이션 및 데이터베이스를 포함하여 다른 IBM 시스템과 완벽하게 통합될 수 있습니다. 또한 IBM Db2 Event Store 에서 캡처한 스트리밍 데이터를 활용할 수도 있습니다.

IBM Integrated Analytics System 의 하드웨어 구성 요소는 다음과 같습니다.

- 3.02GHz 프로세서가 적용된 IBM Power 8 S822L 24 코어 서버 및 통합 IBM 플래시 시스템 (FlashSystem) 900
- 하이브리드 클라우드 또는 논리적 데이터웨어 하우스 구성에 통합되거나 통합되도록 구성할 수 있고 가동 준비가 되어 있는 확장 계층형 스토리지
- 멜라녹스 (Mellanox) 10G 이더넷 스위치 및 브로케이드 (Brocade) SAN 스위치

서버 및 코어의 수와 메모리 및 스토리지의 용량은 사용자의 크기 요구 사항에 따라 달라지며 랙의 3분의 1에서 4개의 랙에 이르는 구성으로 이루어집니다.

이전에 설명된 어플라이언스 형식은 IAS 를 사용할 필요가 없습니다. 필요한 것은 인메모리 BLU 액셀러레이션을 적용한 모든 IBM 플래시 스토리지와 IBM Power 를 사용하는 MPP 아키텍처입니다. IBM 은 이 기술이 다음과 같은 이점을 제공한다고 언급했습니다.

- **성능:** 예기치 않은 응답 시간을 통해 쿼리 성능 저하로 자주 차단되는 "사고의 흐름" 분석이 가능합니다.
- **통합:** Db2 와의 긴밀한 통합으로 모든 애플리케이션에 투명성을 제공합니다.
- **자체 관리 워크로드:** 쿼리가 가장 효율적인 위치에서 실행이 됩니다.
- **투명성:** Db2 에 연결된 애플리케이션은 BLU 액셀러레이션 기능을 전혀 인지하지 못합니다.
- **간소화된 관리:** 어플라이언스의 핸드프리 작업으로 인해 대부분의 데이터베이스 조정 작업이 불필요합니다. IAS 는 System Z와 통합되어 Db2 Z 사용자를 위한 매우 신속한 비즈니스 분석 기능을 제공하는 하이브리드 컴퓨팅 플랫폼을 제공할 수 있습니다. 고성능, 심층 통합, 자체 관리 워크로드 및 간소화된 관리 기능을 갖춘 복잡한 분석을 통한 연결 운영 트랜잭션 처리 기능을 지원합니다. 또한 기능이 투명하여 Db2 Z에 연결된 애플리케이션은 통합을 인지하지 못합니다.

미래 전망

기업은 전문 기술이 없는 사용자가 필요에 따라 직접적으로 관계형 데이터 및 NoSQL 데이터의 연결 조합을 처리하는 데 필요한 분석 시스템을 통합하는 방법을 찾고 있습니다. 이러한 솔루션은 다양한 공급 업체의 클라우드에서 사용할 수 있지만 당분간은 많은 비용과 큰 위험을 들이지 않으면서 시스템을 구현하는 방법을 확보하는 것이 과제가 됩니다. 사용자 데이터 센터 배치와 이에 상응하는 IBM 클라우드 배치가 있는 IBM Integrated Analytics System 과 같은 시스템을 갖추면 이러한 전환이 한층 원활하고 위험성이 줄어듭니다.

과제/기회

이 백서에서 설명한 통합 데이터 수집과 분석 측면을 다루는 많은 기술이 등장하고 있습니다. 적어도 일부 워크로드에 대해서는 IBM Integrated Analytics System 에 대한 합리적인 대안을 제시할 수도 있습니다. 대부분 기능적으로 잘 작동하지만 일반적으로 지원되는 데이터 범위 또는 쉽게 통합될 수 있는 시스템 범위의 측면에서 보면 범위가 제한됩니다. 또한 이 제품이 제공하는 z 시스템의 애플리케이션과 데이터베이스 간의 통합 기능도 거의 제공하지 않습니다.

IBM 은 지원되는 데이터 및 분석 범위, z 시스템과의 통합 능력, 그리고 온프레미스와 IBM 클라우드 사이의 호환성과 연결성을 보여줌으로써 이 제품을 다른 대안과 차별화하고 있습니다.

결론

기업에서 데이터를 수집하여 상관 관계를 분석한 후 조치를 취해야 하는 대량의 데이터를 처리하는 과정에서 직면하는 과제가 현실화되고 있습니다. 기업은 증가하는 데이터 관리 시스템의 수 및 유형과 더불어 데이터 처리를 위한 더욱 복잡한 체계에 대처해야 하며, 이를 분석하여 ML 기반 운영을 통해 데이터를 최대한 활용해서 이익을 얻을 수 있도록 해야 합니다. 이 문제를 해결하기 위한 한 가지 접근법은 모든 기반을 포괄하고 신뢰할 수 있는 기술 공급 업체에 의해 제공되는 데이터 처리, 데이터 구성, 트랜잭션 처리 및 분석을 ML 을 포함한 일관된 단일 시스템을 통해 통합 구성을 마련하여 단순화하는 것입니다. IBM Integrated Analytics System 은 오직 이러한 목적으로 제공됩니다.

상기 설명한 문제에 직면한 기업은 다음 사항을 고려해야 합니다.

- 데이터를 수집, 필터링, 정렬 및 배열하는 데 현재 사용되는 서버 및 스토리지를 포함한 다양한 기능 범위의 인벤토리를 가져옵니다. 이에 따라 데이터를 유용한 구조로 결합하고 데이터에 대한 트랜잭션을 실행하며, 데이터에 대한 분석을 수행하고 데이터로 구동되는 ML 을 지원합니다.
- 일상적인 유지 관리, 패치, 그리고 필요에 따라 시스템에서 시스템으로 데이터를 이동하는 수동 단계를 포함하여 모든 시스템을 관리하는 데 필요한 직원의 작업 시간을 결정합니다.
- 데이터 이동 지연, 데이터 리포맷 및 이동에 소요되는 시간, 인적 실수로 인한 사고, 데이터의 전체 가치를 실질적으로 활용하는 하는 것을 지연시키는 것에 영향을 미치는 모든 경우의 비즈니스 비용을 계산합니다.
- 데이터 수집, 전환, 트랜잭션 처리, 분석, ML 등의 다양한 요구 사항을 처리하는 통합된 요소 시스템으로 이동함으로써 얼마를 저장할 수 있는지, 얼마나 더 많이 얻을 수 있는지 고려하시기 바랍니다. 이러한 계산은 장비 비용, 직원 시간 및 지연으로 인한 기회 비용으로 계산할 수 있어야 합니다.
- IBM Integrated Analytics System 을 해당 목록에 포함시켜 이러한 문제를 해결하는 통합 데이터 관리 및 분석 시스템의 가용성을 살피시기 바랍니다.

IDC 정보

IDC(International Data Corporation)는 IT, 통신 및 소비자 기술 시장을 위한 시장 정보, 자문 서비스 및 이벤트를 최고 수준으로 전세계에 제공하는 회사입니다. IDC 는 IT 전문가, 비즈니스 경영진 및 투자 업계가 기술 구매 및 비즈니스 전략을 사실에 근거한 결정을 할 수 있도록 돕습니다. 1,100 명 이상의 IDC 분석가들이 전세계 110 개 국가에서 기술과 산업 기회 및 경향에 관한 글로벌, 지역 및 지방의 전문 지식을 제공합니다. 50 년 동안 IDC 는 자사 고객들이 주요 비즈니스 목표를 달성하기 위한 전략적 통찰력을 제공해 왔습니다. IDC 는 세계 유수의 기술 미디어, 연구 및 이벤트 회사인 IDG 의 자회사입니다.

글로벌 본사

5 Speen Street
Framingham, MA01701
USA
508.872.8200
Twitter: @IDC
idc-community.com
www.idc.com

저작권 고지

IDC 정보 및 데이터 외부 출판 — IDC 정보를 광고, 보도 자료, 프로모션 자료에 사용하려면 먼저 IDC 부사장 또는 지사장의 사전 서면 승인을 받아야 합니다. 그러한 요청을 할 경우 제안서 초안을 첨부해야 합니다. IDC 는 어떠한 이유로든 외부 사용 승인을 거부할 권리를 갖습니다.

저작권 2018 IDC. 사전 서면 승인 없이 복사하는 것을 금지합니다.

